

## Isolated sequences from the linked *Myf-5* and *MRF4* genes drive distinct patterns of muscle-specific expression in transgenic mice

Ardem Patapoutian<sup>1,\*</sup>, Jeffrey H. Miner<sup>1,\*†</sup>, Gary E. Lyons<sup>2</sup> and Barbara Wold<sup>1,‡</sup>

<sup>1</sup>Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup>Department of Anatomy, University of Wisconsin Medical School, 1300 University Avenue, Madison, WI 53706, USA

\*These two authors contributed equally to this work

†Present address: Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO 63110, USA

‡To whom correspondence should be addressed

### SUMMARY

In developing mouse embryos, *MyoD* family regulatory genes are expressed specifically in muscle precursors and mature myofibers. This pattern, taken together with the well-established ability of *MyoD* family members to convert a variety of cell types to skeletal muscle, suggests a significant role for these genes in regulating skeletal myogenesis. The possibility that expression of these genes may be causally associated with segregation of the myogenic lineage from other mesodermal derivatives, or with the subsequent maintenance of muscle phenotypes at later times, raises the issue of how *MyoD* family genes are themselves regulated during development. In this work, we have initiated studies to identify DNA sequences that govern *Myf-5* and *MRF4* (*herculin*, *myf-6*) transcription. *Myf-5* is the first of the *MyoD* family to be expressed in the developing mouse embryo, while *MRF4* is the most abundantly expressed myogenic factor in postnatal animals. In spite of their strikingly divergent patterns of expression, *Myf-5* and *MRF4* are tightly linked in the mouse genome; their translational start codons are only 8.5 kilobases apart. Here, the 5' flanking regions of the mouse *Myf-5* and *MRF4* genes were separately linked to a bacterial  $\beta$ -galactosidase (*lacZ*) gene, and these constructs were each used to pro-

duce several lines of transgenic mice. Transgene expression was monitored by X-gal staining of whole embryos and by in situ hybridization of embryo sections. For the *Myf-5/lacZ* lines, the most intense transgene expression was in the visceral arches and their craniofacial muscle derivatives, beginning at day 8.75 post coitum (p.c.). This correlates with endogenous *Myf-5* expression in visceral arches. However, while *Myf-5* is also expressed in somites starting at day 8 p.c., transgene expression in the trunk is not observed until day 12 p.c. Thus, the *Myf-5/lacZ* construct responds to early *Myf-5* activators in the visceral arches but not in the somites, suggesting that myogenic determination in the nonsomitic head mesoderm may be under separate control from that of the somitic trunk mesoderm. *MRF4/lacZ* lines displayed an entirely different pattern from *Myf-5*. Transgene expression appeared in muscles starting at day 16.5 p.c. and became increasingly prominent at later times. However, an early wave of myotomal expression that is characteristic of the endogenous *MRF4* was not recapitulated by the transgene.

Key words: myogenesis, *Myf-5*, *MRF4*, transgenic mice, visceral arches, *lacZ*, enhancer trap

### INTRODUCTION

Skeletal muscle is one of many derivatives of mesoderm in vertebrates. Muscles of the head originate from prechordal and paraxial mesodermal cells (Noden, 1991; Couly et al., 1992), most of which populate the visceral arches and then migrate into the developing head. Muscles of the trunk and limbs arise from the somites, which are segmental blocks of paraxial mesoderm that form in pairs on either side of the neural tube (Lyons and Buckingham, 1992; Ordahl and Le Douarin, 1992).

An important contribution to the current view of skeletal myogenesis came from the cloning of *MyoD* (Davis et

al., 1987; Tapscott et al., 1988) and its three close relatives, *myogenin* (Wright et al., 1989; Edmondson and Olson, 1989), *Myf-5* (Braun et al., 1989) and *MRF4/herculin/Myf-6* (Rhodes and Konieczny, 1989; Miner and Wold, 1990; Braun et al., 1990). These genes encode transcription factors of the basic-helix-loop-helix (B-HLH) family (Murre et al., 1989) that bind in vitro to consensus 'E-box' recognition sites. These sites are prominent and functionally significant in many genes expressed specifically in differentiated muscle (Murre et al., 1989; Weintraub et al., 1991). Forced expression of any of the *MyoD* family genes in a variety of nonmuscle cultured cells can convert them to a skeletal muscle phenotype, suggesting that these regulators

play a significant role in determination and differentiation of skeletal muscle (Olson, 1990; Weintraub et al., 1991). The ability of ectopic MyoD and/or Myf-5 to activate skeletal muscle-specific genes in developing *Xenopus* embryos (Hopwood and Gurdon, 1990; Hopwood et al., 1991) and in the hearts of transgenic mice (Miner et al., 1992) supports this view.

Analysis of the expression of these myogenic regulatory genes in several cultured skeletal muscle cell lines has revealed that proliferating myoblasts, which are determined to form muscle, express MyoD (MM14 [Mueller and Wold, 1989]), Myf-5 (L6, BC3H1 [Braun et al., 1989; Mueller and Wold, 1989]), or both (C2C12 [Braun et al., 1989; Miner and Wold, 1990]), while their differentiated counterparts express myogenin always (Emerson, 1990) and MRF4 sometimes (L6J1-C, C2C12 [Rhodes and Konieczny, 1989; Miner and Wold, 1990]). However, it is not at all clear how these few established cell lines are related to muscle and its progenitors in the animal. In situ hybridization experiments on mouse embryo sections have shown that there is a complex pattern of sequential accumulation and disappearance of MyoD family RNAs in developing muscle, and that the pattern varies among different skeletal muscle groups (for review see Buckingham, 1992). Myf-5, the earliest marker of muscle, is first detected at 8 days p.c. in anterior somites, just before the myotome can be recognized as distinct from sclerotome and dermatome by cytological criteria (Ott et al., 1991). Similar expression is sequentially observed in more posterior somites as these structures form in their characteristic rostral-caudal sequence. Myf-5 RNA is subsequently detected at day 9.25 p.c. in the hyoid arch and is followed at day 10 p.c. in the mandibular arch and at day 10.5 p.c. in the forelimb bud. Myogenin and MyoD are expressed in all muscle masses following Myf-5 activation. In contrast, MRF4 is never detected in visceral arches or limb buds by in situ hybridization. However, a wave of transient MRF4 RNA expression is observed in myotomes between days 9 and 11.5 p.c. (Bober et al., 1991; Hinterberger et al., 1991). Late in development, Myf-5 expression declines and becomes undetectable by day 14 p.c. and remains that way (Ott et al., 1991), while MRF4 reappears at day 16 p.c. in all fetal skeletal muscles and becomes the predominant MyoD family regulatory gene expressed in the adult (Rhodes and Konieczny, 1989; Miner and Wold, 1990; Bober et al., 1991; Hinterberger et al., 1991).

The basis for the complex and stringently controlled pattern of differential expression of the MyoD family of regulators is not well understood. Studies of genomic regulatory elements have mainly concentrated on the *myogenin* and *MyoD* genes. Remarkably, only 200 base pairs of proximal 5' flanking sequence from the *myogenin* gene were sufficient to direct myocyte-specific expression (Salminen et al., 1991). The emerging picture for human *MyoD* is quite different. Its 5' flank has been tested for regulatory activity in both cell culture and in transgenic mice. Experiments in cultured cells led to the identification of a region positioned between 18 and 22 kb upstream of the coding region which enhanced transcription from the proximal *MyoD* promoter in 23A2 myoblasts but also, surprisingly, in their nonmyogenic parental cell line C3H 10T1/2, where endogenous

MyoD is not normally expressed. In spite of the apparently inappropriate expression in nonmyogenic cultured cells, this enhancer element drove expression in a muscle-restricted pattern in transgenic mice (Goldhamer et al., 1992).

In contrast to *MyoD* and *myogenin*, little is presently known about the regulation of MRF4 and Myf-5. Myf-5 transcripts appear before any of the other *MyoD* family RNAs (Ott et al., 1991). MRF4 is most notably expressed in late fetal and postnatal muscle where it quantitatively predominates over the other *MyoD* family transcripts (Miner and Wold, 1990), suggesting a role for MRF4 in maintenance of differentiated muscle. Furthermore, the close physical linkage of Myf-5 and MRF4, presumably the result of an ancient gene duplication, presents an interesting problem in the evolution, organization and utilization of regulatory elements. The two genes share a stringent specificity for expression in skeletal muscle, but show highly disparate developmental regulation. In preliminary experiments, we found expression of reporter genes carrying Myf-5 and MRF4 flanking sequences in cultured cell lines to be minimal. To overcome this assay limitation and to gain access to the full developmental diversity of MRF4 and Myf-5 expressing cells in the animal, we have produced transgenic mice expressing the reporter constructs. These experiments have allowed us to identify sequence elements from Myf-5 and MRF4 that specify expression in their distinct, spatially and temporally restricted patterns.

## MATERIALS AND METHODS

### Construction of the transgenes

5.5 kb of the 5' flanking region of the mouse Myf-5 gene were isolated as a *Bam*HI to *Sac*I fragment from the original mouse Myf-5/*herculin* phage described previously (Miner and Wold, 1990). This fragment contains the Myf-5 proximal promoter and putative transcription start site (Ott et al., 1991). A bacterial  $\beta$ -galactosidase gene containing an introduced consensus Kozak translation initiation sequence was placed downstream of this in Bluescript II KS+ (Stratagene) to make MYF5Z.

The original Myf-5/*herculin* phage contained only about 350 base pairs of *herculin* (MRF4) 5' flanking DNA. Genomic Southern blots (J. M., unpublished) had indicated the existence of another *Bam*HI site approximately 10 kb upstream of the previously identified site (Miner and Wold, 1990), or 6.5 kb upstream of the MRF4 coding region. We cloned this 10 kb *Bam*HI fragment by constructing a phage lambda library from ~9-11 kb *Bam*HI fragments of mouse genomic DNA (electroeluted from an agarose gel) using the vector lambda gem-12 (Promega). The library was probed with the MRF4 transcribed region, and the insert of a positive phage was subcloned into Bluescript II KS+. Restriction analysis showed that this phage contained the MRF4 coding region as well as 6.5 kb of 5' flanking DNA, as expected. To be sure of including the MRF4 transcription start site in the *lacZ* construction, we used the ~6.5 kb *Bam*HI to *Sal*I fragment as a foundation and then ligated 52 base pairs of additional contiguous sequence, synthesized as two complementary oligonucleotides by the Caltech Microchemical Facility. This sequence was added because preliminary primer extension assays (J. M., unpublished) had indicated that it should contain the major transcription start site. Therefore, the 3' end of the mouse MRF4 sequences used here is nucleotide 64 of the reported *herculin* genomic sequence (Miner and Wold, 1990).

### Preparation of DNA and production of transgenic mice

MYF5Z and MRF4Z plasmids were purified by cesium chloride density gradient ultracentrifugation. To liberate the constructs from vector sequences, MYF5Z was cut with *Kpn*I, *Bam*HI and *Sca*I, and MRF4Z was cut with *Bam*HI and *Xho*I. To isolate the fragments for microinjection, the restriction digests were loaded onto 10–40% preformed sucrose gradients and ultracentrifuged in a SW41 rotor at 26,000 revs/minute for 24 hours (Maniatis et al., 1982). 300 µl fractions were collected and analyzed by agarose gel electrophoresis. DNA was ethanol-precipitated out of the appropriate fractions and dissolved in 10 mM Tris (pH=7.5), 0.1 mM EDTA (pH=8). Transgenic mice were produced by pronuclear microinjection of single cell mouse embryos from a (C57BL/6×DBA/2)F<sub>1</sub>×(C57BL/6×DBA/2)F<sub>1</sub> cross as described (Hogan et al., 1986; Miner et al., 1992).

### Analysis of transgenic mice

Transgenic founder mice were identified by Southern blot or polymerase chain reaction analysis of tail DNA. Male founders and male offspring of female founders were mated in most cases with C57BL/6×DBA/2 hybrid females, though sometimes the parental inbred strains were used. For postnatal analyses, pups of various ages were killed, skinned, fixed in 4% formaldehyde in phosphate-buffered saline (PBS) and stained in PBS containing 35 mM potassium ferricyanide, 35 mM potassium ferrocyanide, 1.5 mM magnesium sulfate, and 1 mg/ml X-gal (US Biochemicals Corp.) overnight at 37°C. For prenatal analyses, timed pregnant females were killed at the desired day of gestation, and embryos were dissected out of the uterine horns into PBS. They were fixed and stained as above, though for older embryos the staining solution was supplemented to contain 0.2% sodium deoxycholate and 0.1% Nonidet P-40 to enhance X-gal penetration.

### In situ hybridization

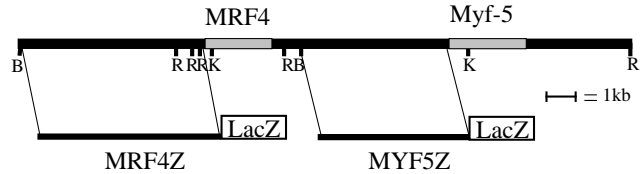
In situ hybridization was performed on 5 to 7 µm paraffin sections. The procedures used for section treatment, hybridization and washings are described by Lyons et al. (1990). Hybridizations were carried out at 50°C for ~16 hours in 50% deionized formamide, 0.3 M NaCl, 20 mM Tris-HCl (pH 7.4), 5 mM EDTA, 10 mM sodium phosphate (pH 8), 10% dextran sulfate, 1× Denhardt's solution, 50 µg/ml of yeast RNA, with 50–75,000 cts/minute/µl of cRNA labeled with <sup>35</sup>S-labeled UTP (>1000 Ci/mmol, Amersham). Washing was at 65°C in 50% formamide, 2× SSC, and 10 mM DTT. Slides were then treated with RNase A (20 µg/ml) (Boehringer Mannheim) for 30 minutes at 37°C. After washes, slides were processed for standard autoradiography with Kodak NTB-2 nuclear track emulsion and exposed for 7 days. Analysis was carried out with both light- and dark-field optics on a Zeiss Axiophot microscope.

For the *Myf-5* probe, a 310-bp *Bal*I-*Apa*I fragment of the first exon of the mouse *Myf-5* gene was used as described by Ott et al. (1991). For the *MRF4* probe, the 680 bp 3' *Pst*I fragment of the mouse gene was used as described by Bober et al. (1991). For the *lacZ* probe, the 3 kb *Bss*III fragment of the MRF4Z construct, which also contains the SV40 polyadenylation signal sequence, was used.

## RESULTS

### Production of transgenic mice

5.5 kb of *Myf-5* and 6.5 kb of *MRF4* 5' flanking regions were each linked to the bacterial  $\beta$ -galactosidase gene. These constructs (Fig. 1) are called MYF5Z and MRF4Z,



**Fig. 1.** Structure of the mouse *MRF4/Myf-5* locus and the respective transgenes. The stippled areas approximate transcribed regions, but for *Myf-5* only the location of the first exon is known. *lacZ* segments are not drawn to scale. See Materials and Methods for construction details. B, *Bam*HI; R, *Eco*RI; K, *Kpn*I.

respectively. Multiple transgenic lines were produced by pronuclear microinjection. All four DNA-positive lines produced with the MYF5Z construct expressed the transgene; two of the four lines, MYF5Z-21 and MYF5Z-29, showed the most intense  $\beta$ -galactosidase ( $\beta$ -gal) staining and hence were more comprehensively characterized. The two other lines, MYF5Z-9 and MYF5Z-46, showed an overlapping but reduced expression pattern. Seven MRF4Z transgenic lines were produced: three (MRF4Z-28, -45 and -49) expressed the transgene in a muscle-specific manner, three did not express the reporter gene at all, and a single line, called MRF4Z-4, exhibited ectopic expression in several diverse tissues characteristic of strong, site-of-insertion position effects (Allen et al., 1988; Gossler et al., 1989).

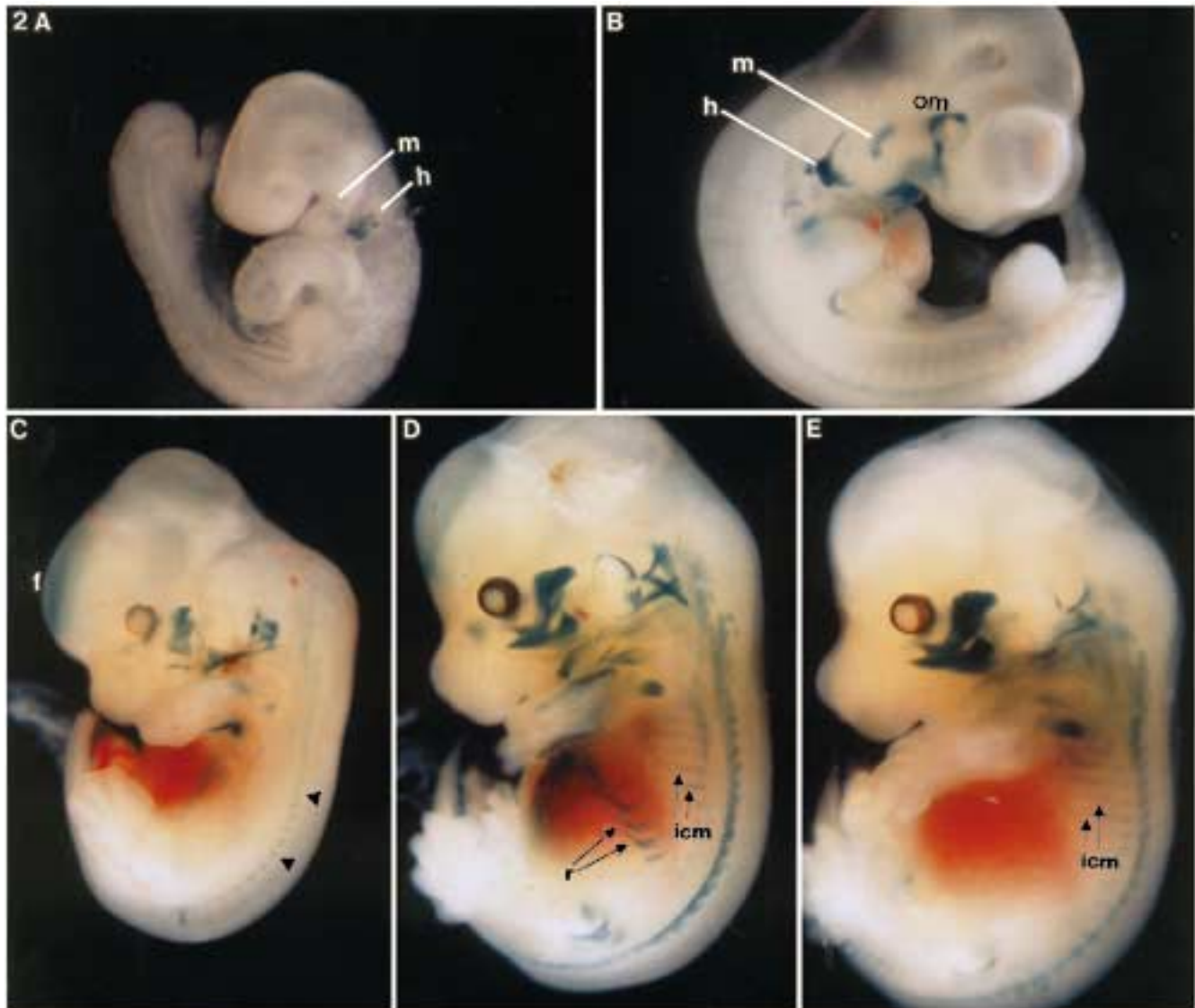
### MYF5Z transgenic mice

To survey developmental expression patterns of the transgene in detail, we used histochemical staining for  $\beta$ -gal beginning at day 8.75 p.c. (E8.75) and continuing through birth. This assay is highly sensitive and parallel controls showed all staining, except inside the gut of postnatal pups, to be specified by the activity of the transgene. The E8.75 (15 somites) embryos examined from MYF5Z-21 and -29, the two best-expressing lines, contained  $\beta$ -gal-positive cells in the hyoid and mandibular arches but not in somites (Fig. 2A). Since prior in situ hybridizations had shown that *Myf-5* transcripts begin to accumulate in somites at E8.0 and in visceral arches at E9.25 (Ott et al., 1991), it appears that this construct contains sequences sufficient to specify the early visceral arch expression, but insufficient to drive early somitic expression. As development proceeded, both the number of cells expressing *lacZ* and the intensity of expression increased. At E10.5, the hyoid arch exhibited very intense expression, and  $\beta$ -gal-positive cells were visible in the developing ocular muscles (Fig. 2B). At E12, myotomal staining became detectable, then increased in intensity at E12.5 (Fig. 2C). Also at E12.5, staining was visible in the developing muscles of the head which are derived from the visceral arches and in the muscle masses of the proximal forelimb. When myotomal expression of the transgene becomes detectable, it already encompasses the full length of the embryo (Fig. 2C,D), so the anterior-posterior sequence of endogenous *Myf-5* activation that corresponds to early myotome segregation is not recapitulated by the transgene. Transgene expression was most prominent at E13.5 with the appearance of  $\beta$ -gal-positive cells in the

intercostal muscles and developing abdominal muscles (Fig. 2D). Beginning at day 14, when endogenous *Myf-5* RNA levels begin to decrease (Ott et al., 1991), muscle-specific *lacZ* expression declined and became undetectable shortly after birth (data not shown). The expression patterns described were present in at least two different transgenic lines (Fig. 2D,E), confirming that the transgene expression in these tissues resulted from regulatory elements present in the MYF5Z construct, not from site-of-integration position effects. This  $\beta$ -gal pattern reflects a substantial subset

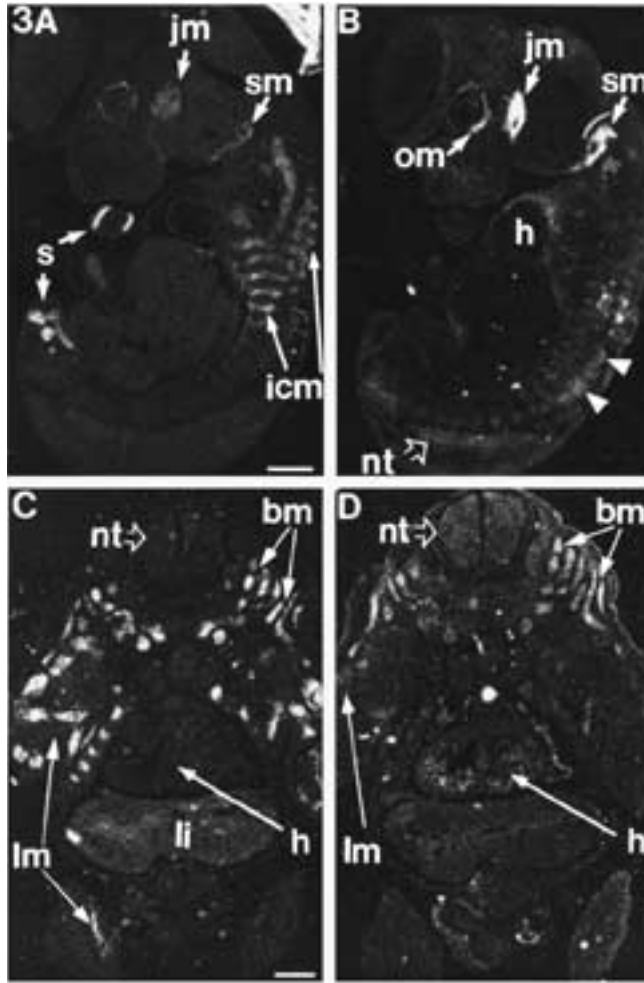
of the endogenous *Myf-5* expression pattern as described previously by in situ hybridization (Ott et al., 1991).

To compare endogenous and transgene expression directly, and to verify the sensitivity of the whole-mount embryo X-gal staining assay, E11.5 and E12.5 embryos from MYF5Z-21 and MYF5Z-29 transgenic lines were sectioned, and in situ hybridization experiments with *Myf-5* and *lacZ* probes were performed on adjacent sections. The *lacZ* expression pattern observed in these experiments was identical to the whole-embryo X-gal staining pattern. In



**Fig. 2.** Whole-mount histochemical staining of MYF5Z embryos at various developmental stages. (A) A 15 somite (E8.75) MYF5Z-21 embryo. The hyoid arch and the mandibular arch are positive for *lacZ* transcripts. (B) E10.5 MYF5Z-21 embryo. The hyoid arch is the most intensely stained structure and staining is also visible in the mandibular arch and in the ocular muscles. Ectopic staining specific to MYF5Z-21 embryos is observed in the nervous system as well as in the epithelium of the forelimb-bud. (C) E12.5 MYF5Z-21 embryo. Staining is evident in muscular derivatives of the visceral arches, in the ocular muscles, in developing proximal forelimb muscles, and in the segmented myotomes. Additional ectopic expression is observed in the forebrain. (D) E13.5 MYF5Z-21 embryo. Staining is similar to that observed in C, with additional muscle-specific staining in intercostal, proximal hindlimb, and dorsal neck and head muscles. Ectopic *lacZ* expression is observed in the ribs. (E) E13.5 MYF5Z-29 embryo. Muscle-specific staining is identical to that observed in D, indicating that this staining is due to regulatory sequences in the transgene and not those found at the site of integration. f, forebrain, h, hyoid arch, icm, intercostal muscles, m, mandibular arch, om, ocular muscles, r, rib, arrowhead, myotome.

most cases, *Myf-5* and *lacZ* RNA distribution was similar, with *lacZ* representing a subset of the endogenous *Myf-5* expression. At E11.5 (Fig. 3A,B), head muscles were positive with both *Myf-5* and *lacZ* probes, while somitic expression was observed only with the *Myf-5* probe. At E12.5 (Fig. 3C,D), trunk muscles as well as the proximal forelimb muscles expressed the transgene.



**Fig. 3.** Expression of *Myf-5* and *lacZ* in MYF5Z transgenic embryos as detected by in situ hybridization. (A,B) Parasagittal sections of E11.5 MYF5Z-21 embryo hybridized to *Myf-5* (A) and *lacZ* (B) probes. Both *Myf-5* and *lacZ* transcripts are expressed in jaw (jm) and shoulder (sm) muscles, while only *Myf-5* shows significant hybridization in intercostal muscles (icm) and tail somites (s) at this stage. While *lacZ* is expressed in ocular muscles (om), these muscles are not present in the more lateral section (A) probed with *Myf-5*. Notice that the pigment layer of the retina refracts under dark-field illumination. Some ectopic expression of *lacZ* is observed in the neural tube (nt), dorsal ganglia (arrowheads) and surrounding the heart. (C,D) Frontal sections of an E12.5 MYF5Z-29 embryo hybridized to *Myf-5* (C) and *lacZ* (D) probes. Both *Myf-5* and *lacZ* transcripts are expressed in the back (bm) and limb muscles (lm), with limb muscles showing more intense signal with the *Myf-5* probe than with the *lacZ* probe. The transgene also shows some faint ectopic expression in the heart (h) and neural tube (nt) which was not observed in the whole-embryo X-gal assay. li, liver. Scale bar = 500  $\mu$ m.

In individual MYF5Z transgenic lines, *lacZ* was expressed ectopically in a few non-skeletal muscle tissues (Fig. 2B-D, 3B,D), as observed by both whole-mount staining of embryos with X-gal and in situ hybridization experiments. Since these expression patterns were each characteristic of one line only, and not shared by any of the other three transgenic lines (Fig. 2D,E), we suggest that the observed ectopic expression is probably due to enhancer elements trapped near the insertion site of the construct (Lacy et al., 1983; Kothary et al., 1988; Al-Shawi et al., 1990).

### MRF4Z transgenic mice

Three MRF4Z lines expressed the *lacZ* transgene in multiple muscle groups beginning at E16 (data not shown). Expression of the transgene increased throughout the perinatal period (Fig. 4) in parallel with the observed pattern of endogenous MRF4 expression (Bober et al., 1991; Hinterberger et al., 1991). However, MRF4Z transgenic mice did not express the transgene transiently in embryonic myotomes as the endogenous *MRF4* gene is expressed.

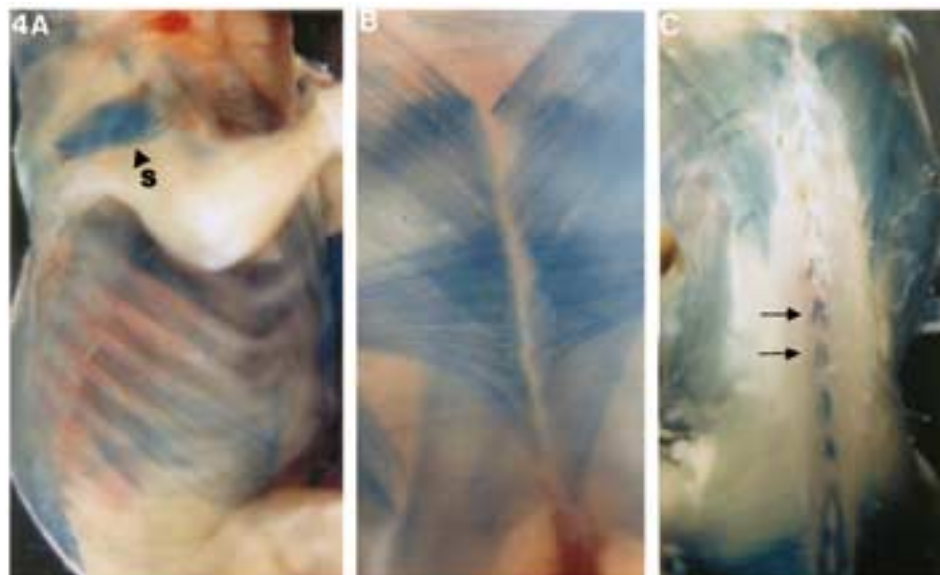
Some muscles expressed the MRF4Z transgene more intensely than others. For example, muscles of the limbs showed faint  $\beta$ -gal staining (Fig. 4A), whereas the spinodeltoideus muscle (Fig. 4A), levator auris longus muscle (Fig. 4B), and intervertebral muscles (Fig. 4C) expressed *lacZ* particularly strongly. To compare the spatial distribution of the endogenous *MRF4* and *lacZ* transcripts, abdominal muscles were dissected and in situ hybridization reactions were performed with *MRF4* and *lacZ* probes on adjacent sections (Fig. 5). Both *MRF4* and *lacZ* transcripts were detected in all muscle fibers, but showed no activity in the nearby connective tissue.

Because three of four independent transgenic MRF4Z lines exhibited the same general patterns of *lacZ* expression, and because this pattern parallels that of endogenous MRF4, we attribute this expression to regulatory sequences present in the transgene. However, one MRF4Z line, MRF4Z-4, expressed *lacZ* in a different and unexpected fashion (Fig. 6). With the exception of the spinotrapezius muscle, which expressed intensely starting at E16, MRF4Z-4 transgene expression did not correlate with the endogenous MRF4 pattern (Bober et al., 1991; Hinterberger et al., 1991) or with the other MRF4Z lines. Instead, this line expressed *lacZ* in a variety of other tissues at different developmental time periods including vibrissae (Fig. 6A), hand pads, olfactory bulbs, knee and elbow joints (Fig. 6A) and mid-brain. This diversity suggests a dominant position effect coupled, perhaps, with damage to one or more copies of the transgene.

### DISCUSSION

The MyoD family of regulators have been inferred to play a significant role in skeletal myogenesis. Three lines of evidence support this idea. First, all four MyoD family members can recruit diverse cultured nonmuscle cells to a skeletal muscle phenotype (Olson, 1990; Weintraub et al., 1991), and both MyoD and Myf-5 have been shown to be capable of activating skeletal muscle genes ectopically in devel-

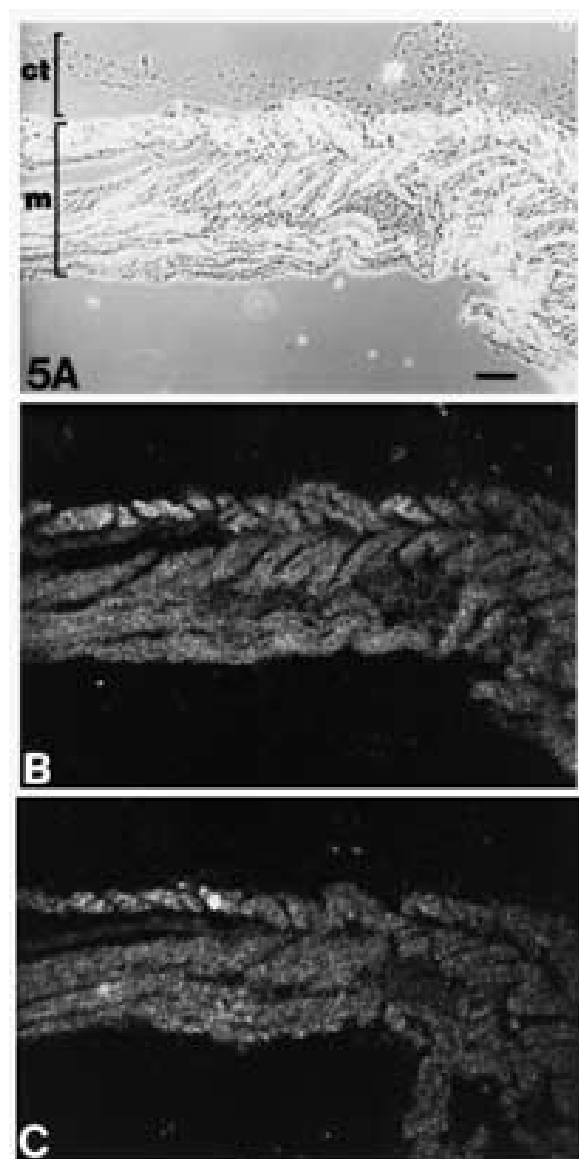




**Fig. 4.** Whole-mount X-gal histochemical staining of MRF4Z postnatal mice. (A) A lateral view of the trunk of a 2-day-old MRF4Z-28 mouse pup. Most muscles of the trunk contain fibers that are positive for *lacZ* activity, including latissimus dorsi, pectoralis and the abdominals. The spinodeltoideus muscle (s) is consistently the most intensely stained in MRF4Z transgenics; the limbs exhibit considerably less *lacZ* activity. (B) An enlarged dorsal view of the same pup showing staining in muscles of the head and the neck. (C) A dorsal view of the trunk of a 7-day-old MRF4Z-45 pup. Most muscles are  $\beta$ -gal positive; staining is striking in the paired intervertebral muscles (arrowheads) along the anterior-posterior axis. The top is anterior in all three cases.

oping embryos (Hopwood and Gurdon, 1990; Hopwood et al., 1991; Miner et al., 1992). This argues for a direct or indirect role for the MyoD family in the myogenic determination process. Second, molecular studies suggest a direct interaction between MyoD family regulators and transcriptional enhancers of many terminally differentiated muscle-specific genes (Olson, 1990; Weintraub et al., 1991). This is consistent with a direct and ongoing role in execution and maintenance of terminal muscle differentiation. Finally, the MyoD family of transcription factors (most notably Myf-5) are expressed early in premuscle cells (Sassoon et al., 1989; Ott et al., 1991; Bober et al., 1991; Hinterberger et al., 1991), which reinforces a role for these regulators in myogenic determination prior to overt differentiation of muscle. However, little is known about the molecular and cellular processes that govern the segregation of the myogenic lineage from other mesodermal derivatives in the developing mouse embryo. If the activities of these MyoD family regulatory proteins are partly or wholly responsible for establishing the skeletal muscle phenotype, then learning how expression of these regulators is initiated, maintained, and terminated during embryogenesis is crucial for understanding how myogenic cell fate is specified and executed. Apart from the postulated importance of *MyoD* family products in regulating myogenesis, their complex and diverse patterns of expression during development make them very useful markers for distinct subdivisions of myogenic populations.

Here we have identified DNA sequences from the *Myf-5/MRF4* locus that direct expression of the bacterial  $\beta$ -



**Fig. 5.** *MRF4* and *lacZ* transgene expression in abdominal muscle sections of a 6.5-day-old MRF4Z-45 transgenic pup as detected by in situ hybridization. (A) A phase-contrast micrograph of a section showing abdominal muscle fibers (m) associated with some adjacent connective tissue (ct). (B) Dark-field micrograph of A hybridized with the *lacZ* probe. (C) Dark-field micrograph of an adjacent section hybridized with the *MRF4* probe. All muscle fibers are positive with both probes, while the connective tissue associated with the muscle is negative. Scale bar = 100  $\mu$ m.



**Fig. 6.** Whole-mount histochemical staining of the MRF4Z-4 transgenic mice demonstrates presumed insertion site-dependent expression specific to this transgenic line. (A) E15.5 embryo. Staining is evident in the vibrissae (whisker pads), around the opening of the mouth, and in connective tissue of the elbow and the knee. The only muscle staining observed in this line is in the spinotrapezius muscle of the back. (B) A dorsal view of the trunk of a 2-day-old pup showing intense staining of the spinotrapezius muscle of the back. Staining in elbow and knee connective tissue persists.

*galactosidase* gene in patterns reflecting that of endogenous Myf-5 and MRF4. The MYF5Z transgenic mice, which carried 5.5 kb of upstream *Myf-5* sequences, expressed the *lacZ* reporter gene in a large subset of the cells which express endogenous *Myf-5* RNA. Myf-5, the earliest known marker for myogenic precursors, is initially expressed in a group of cells in somites and visceral arches which later become muscle cells of the trunk and head, respectively (Ott et al., 1991). Our results suggest that visceral arch expression of Myf-5 is under separate control from early Myf-5 expression in the somite, since we have shown that *Myf-5* visceral arch regulation is recapitulated in our construct, but early somitic expression is not. Although confirmation will require isolation of sequences that do direct early somitic expression, the clear implication is that there are distinct regulatory pathways by which *Myf-5* expression, and perhaps muscle determination, is initiated in somitic trunk versus unsegmented head mesoderm. In a similar vein, expression of *Myf-5* observed in our experiments subdivides somitic expression by time (early versus late) and by the subset of somitic Myf-5-positive cells that also show *lacZ* expression by in situ hybridization. In addition, the failure of this MYF5Z construct to be expressed when transfected into C2C12 cells (A.P., unpublished), which do express some endogenous Myf-5 (Miner and Wold, 1990), is consistent with the organization of *Myf-5* control regions into multiple segregated regulatory elements, and it is likely that only a subset of these are included in this construct.

The MRF4Z transgenic mice expressed the transgene in muscle cells of embryos starting at E16.5, and the intensity of  $\beta$ -gal staining increased through the first week of postnatal development. This general expression pattern reflects that of endogenous MRF4 as determined by in situ hybridization and northern blot experiments (Bober et al., 1991; Hinterberger et al., 1991). However, endogenous MRF4 also displays a separate, transient expression pattern in myotome between E9 and 11.5 (Bober et al., 1991; Hinterberger et al., 1991), but we did not detect any somitic expression of the transgene in MRF4Z lines. Interestingly,

this construct, when transfected into C2C12 cells, which express some endogenous MRF4 upon differentiation (Miner and Wold, 1990), is active only in the presence of high concentrations of cotransfected MyoD family regulatory factors (A.P., unpublished). Thus, for both *MRF4* and *Myf-5*, sequences that normally regulate early myotomal as well as C2C12 transcription were apparently not included in the constructs. Of course there are additional mechanisms of regulation that may contribute to differences between transgene expression and endogenous RNA levels, including specific methylation of sequences within the transgene, or post-transcriptional mechanisms such as RNA stability. It is also possible that all regulatory regions are present in our constructs but are dependent on the linked state of *Myf-5* and *MRF4*.

In these experiments, some ectopic  $\beta$ -galactosidase was observed in individual lines and was not reproduced in other transgenic lines with the same construct. This raises the general issue, crucial in experiments of this design, of proper attribution of transgene expression patterns. Two criteria were applied to assign expression to the elements from *MRF4* and *Myf-5*. First, for any pattern attributed to the construct, transgene expression was observed in independent transgenic lines carrying the same construct, arguing that the signal was due to the construct itself and not to position effects at the site of insertion. The second criterion was to compare directly the transgene expression pattern with that of the endogenous gene by in situ hybridization. In this study, all elements of transgene expression that were shared by multiple lines of transgenic mice from the same construct also correlated with RNA expression from their respective endogenous genes, supporting the idea that the elements used in these experiments represent true regulatory regions of *Myf-5* and *MRF4*. In situ hybridization experiments were especially informative in the MRF4Z pups where penetrance of histochemical reagents into more mature muscle became a technical limitation. The line-specific ectopic transgene expression patterns were most likely due to adventitious enhancer trapping, in which the con-

struct integrates near one or more endogenous, active genes and the site of integration exerts some regulatory influence over the transgene (Lacy et al., 1983; Kothary et al., 1988; Al-Shawi et al., 1990). Three of 11 lines displayed position effect phenomena at some stage of development, with two other lines, MYF5Z-21 and MYF5Z-29, showing some ectopic expression together with the expected pattern. In the most extreme case, MRF4Z-4, the typical MRF4 pattern was almost entirely suppressed and a wide array of ectopic sites were substituted. Curiously expression in just one major muscle, the spinotrapezius (Fig. 6), was very prominent, and correlated with the temporal regulation of the endogenous MRF4.

Of the isolated regulatory elements, perhaps the most interesting is the *Myf-5* 5' flank, which responds to early signals in the developing head; it will be interesting to more narrowly define the sequences responsible. Since craniofacial muscles are derived from unsegmented paraxial mesoderm (Couly et al., 1992) and not from the somitic mesoderm that gives rise to trunk and limb muscles, a regulatory element that responds to early myogenic signals in the head but not in the somites may ultimately permit molecular-level description of how these two myogenic lineages are differentially specified. Also, while cell culture experiments have shown that forced expression of any of the four MyoD family regulators can initiate a myogenic pathway, the issue of whether individual regulators are functionally different in vivo remains open and will require both gain-of-function and loss-of-function manipulations. The *cis*-regulatory segments defined here should serve as effective tools for testing the in vivo consequences of specifically altering the expression pattern of myogenic regulators by, for example, expressing *MRF4* in visceral arches and their derivatives.

We are indebted to S. Darling and J. Rossant for the *lacZ* plasmid and H. Arnold for *MRF4* and *Myf-5* cDNAs. We thank S. Fukuda and K. Lee for technical assistance. We also thank D. Anderson, J. Sanes, and members of the Wold group for helpful comments on the manuscript.

A. P. was supported by an NIH Predoctoral Training Grant. This work was supported by grants to B. W. from the NIH and MDA, and to G. L. from MDA.

## REFERENCES

- Allen, N. D., Cran, D. G., Barton, S. C., Hettle, S. and Reik, W. (1988). Transgenes as probes for active chromosomal domains in mouse development. *Nature* **333**, 852-855.
- Al-Shawi, R., Kinnaird, J., Burke, J. and Bishop, J. (1990). Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol. Cell Biol.* **10**, 1192-1198.
- Bober, E., Lyons, G. E., Braun, T., Cossu, G., Buckingham, M. and Arnold, H.-H. (1991). The muscle regulatory gene, *Myf-6*, has a biphasic pattern of expression during early mouse development. *J. Cell Biol.* **113**, 1255-1265.
- Braun, T., Bober, E., Winter, B., Rosenthal, N. and Arnold, H. H. (1990). *Myf-6*, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. *EMBO J.* **9**, 821-831.
- Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. and Arnold, H. H. (1989). A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *EMBO J.* **8**, 701-709.
- Buckingham, M. (1992). Making muscle in mammals. *Trends Gen.* **8**, 144-149.
- Couly, G. F., Coltey, P. M. and Le Douarin, N. M. (1992). The developmental fate of the cephalic mesoderm in quail-chick chimeras. *Development* **114**, 1-15.
- Davis, R. L., Weintraub, H. and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000.
- Edmondson, D. G. and Olson, E. N. (1989). A gene with homology to the *myc* similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes Dev.* **3**, 628-640.
- Emerson, C. P. (1990). Myogenesis and developmental control genes. *Curr. Opinion in Cell Biol.* **2**, 1065-1075.
- Goldhamer, D. J., Faerman, A., Shani, M. and Emerson Jr., C. P. (1992). Regulatory elements that control the lineage-specific expression of *myoD*. *Science* **256**, 538-542.
- Gossler, A., Joyner, A. L., Rossant, J. and Skarnes, W. C. (1989). Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463-465.
- Hinterberger, T. J., Sassoon, D. A., Rhodes, S. J. and Konieczny, S. F. (1991). Expression of the muscle regulatory factor MRF4 during somite and skeletal myofiber development. *Dev. Biol.* **147**, 144-156.
- Hogan, B., Costantini, F. and Lacy, E. (1986). *Manipulating the Mouse Embryo*. New York: Cold Spring Harbor Laboratory.
- Hopwood, N. D. and Gurdon, J. B. (1990). Activation of muscle genes without myogenesis by ectopic expression of MyoD in frog embryo cells. *Nature* **347**, 197-200.
- Hopwood, N. D., Pluck, A. and Gurdon, J. B. (1991). *Xenopus* Myf-5 marks early muscle cells and can activate muscle genes ectopically in early embryos. *Development* **111**, 551-560.
- Kothary, R., Clapoff, S., Brown, A., Campbell, R., Peterson, A. and Rossant, J. (1988). A transgene containing *lacZ* inserted into the *dystonia* locus is expressed in the neural tube. *Nature* **335**, 435-437.
- Lacy, E., Roberts, S., Evans, E. P., Burtenshaw, M. D. and Constantini, F. D. (1983). A foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues. *Cell* **34**, 343-358.
- Lyons, G. E. and Buckingham, M. E. (1992). Developmental regulation of myogenesis in the mouse. *Sem. Dev. Biol.* **3**, 243-253.
- Lyons, G. E., Ontell, M., Cox, R., Sassoon, D. and Buckingham, M. (1990). The expression of myosin genes in developing skeletal muscle in the mouse embryo. *J. Cell Biol.* **111**, 1465-1476.
- Maniatis, T., Fritsch, E. R. and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor.
- Miner, J. H., Miller, J. B. and Wold, B. J. (1992). Skeletal muscle phenotypes initiated by ectopic MyoD in transgenic mouse heart. *Development* **114**, 853-860.
- Miner, J. H. and Wold, B. (1990). Herculin, a fourth member of the *MyoD* family of myogenic regulatory genes. *Proc. Natl. Acad. Sci. USA* **87**, 1089-1093.
- Mueller, P. R. and Wold, B. (1989). *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**, 780-786.
- Murre, C., Schonleber McCaw, P., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H. and Baltimore, D. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537-544.
- Noden, D. M. (1991). Vertebrate craniofacial development: the relation between ontogenetic process and morphological outcome. *Brain Behav. Evol.* **38**, 190-225.
- Olson, E. N. (1990). MyoD family: a paradigm for development? *Genes Dev.* **4**, 2104-2111.
- Ordahl, C. P. and Le Douarin, N. M. (1992). Two myogenic lineages within the developing somite. *Development* **114**, 339-353.
- Ott, M.-O., Bober, E., Lyons, G., Arnold, H. and Buckingham, M. (1991). Early expression of the myogenic regulatory gene, *myf-5*, in precursor cells of skeletal muscle in the mouse embryo. *Development* **111**, 1097-1107.
- Rhodes, S. J. and Konieczny, S. F. (1989). Identification of MRF4: A new member of the muscle regulatory factor gene family. *Genes Dev.* **3**, 2050-2061.
- Salminen, A., Braun, T., Buchberger, A., Jurs, S., Winter, B. and



- Arnold, H.-H.** (1991). Transcription of the muscle regulatory gene *MYF4* is regulated by serum component, peptide growth factors and signaling pathways involving G proteins. *J. Cell Biol.* **115**, 905-917.
- Sassoon, D., Lyons, G., Wright, W. E., Lin, V., Lassar, A., Weintraub, H. and Buckingham, M.** (1989). Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. *Nature (London)* **341**, 303-307.
- Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P.-F., Weintraub, H. and Lassar, A. B.** (1988). MyoD1: a nuclear phosphoprotein requiring a myc homology region to convert fibroblasts to myoblasts. *Science* **242**, 405-411.
- Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benzeira, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y. and Lassar, A.** (1991). The *myoD* gene family: nodal point during specification of the muscle cell lineage. *Science* **251**, 761-766.
- Wright, W. E., Sassoon, D. A. and Lin, V. K.** (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD1. *Cell* **56**, 607-617.

(Accepted 27 January 1993)